

A novel, divergence based, regression for compositional data

M. Tsagris

Independent researcher, Athens, mtsagris@yahoo.gr

Abstract

In compositional data, an observation is a vector with non-negative components which sum to a constant, typically 1. Data of this type arise in many areas, such as geology, archaeology, biology, economics and political science amongst others. The goal of this paper is to propose a new, divergence based, regression modelling technique for compositional data. To do so, a recently proved metric which is a special case of the Jensen-Shannon divergence is employed. A strong advantage of this new regression technique is that zeros are naturally handled. An example with real data and simulation studies are presented and are both compared with the log-ratio based regression suggested by Aitchison in 1986.

Keywords: compositional data, Jensen-Shannon divergence, regression, zero values, ϕ -divergence

1 Introduction

Compositional data is a special type of multivariate data in which the elements of each observation vector are non-negative and sum to a constant, usually taken to be unity.

$$\mathbb{S}^d = \left\{ (x_1, \dots, x_D) | x_i \geq 0, \sum_{i=1}^D x_i = 1 \right\} \quad (d = D - 1)$$

A typical data set with $D = 3$ components has the following form:

$$\begin{bmatrix} 0.775 & 0.195 & 0.030 \\ 0.719 & 0.249 & 0.320 \\ \vdots & \vdots & \vdots \end{bmatrix}$$

When $D = 3$, a good way to visualize compositional data is the so called ternary diagram. Figure 1 presents the Arctic lake data (we will see this data set again later and describe it in more detail), which consist of three components, sand, silt and clay. These data are available in Aitchison (2003, p. 359) or the in R package *compositions* along with more compositional data and techniques to analyze such data. Sand is a naturally occurring granular material composed of finely divided rock and mineral particles. Silt is a granular material of a size somewhere between sand and clay whose mineral origin is quartz and feldspar. As for clay, it is a fine-grained natural rock or soil material that combines one or more clay minerals with traces of metal oxides and organic matter.

The closer a point is to a vertex, the higher its value in that corresponding component. If the point is on the vertex, it means that its value in the component corresponding to that vertex is 1 and the other two values are 0. On the opposite side, the further away a point is from a vertex, the lowest its value in the corresponding component. If a point lies on an edge,

it means, that the value of the value of the component corresponding to that vertex is zero. If the point is on the middle of that edge, the values of the other two components are exactly 0.5. Points at the barycenter indicate that the values are equal (to $1/3$). For example, the points lying at the base of the triangle in Figure 1 have very low values of clay and higher values sand than silt.

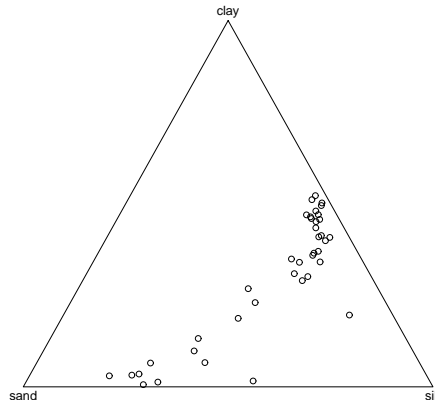


Figure 1: Ternary diagram of the Arctic lake data (Aitchison, 2003).

Analysis of such data may be carried out in various ways which can be summarized in two directions: either by transforming the data or not. In the first direction, the most popular transformations are the log of ratios formed by the components (Aitchison, 1982; 2003). Other approaches include taking the square root of the data, resulting in data which lie on the hypersphere (Stephens, 1982 and Scealy & Welsh, 2011). Another parametric model of this school of thought is the Dirichlet distribution (Gueorguieva et al., 2008). In all of these cases, regression models have been developed.

An important issue in compositional data is the presence of zeros, which cause problems for the logarithmic transformation. The issue of zero values in some components is not addressed in most papers, but see Neocleous et al.(2011) for an example of discrimination in the presence of zeros. Only when treated as directional data, compositional data have no problem with zeros. Log-ratio and Dirichlet regression zero imputation techniques must be applied prior to fitting the models.

The paper proposes a novel regression method based on a recently suggested metric. It is a new metric for probability distributions (Endres & Schindelin, 2003 and Österreicher & Vajda, 2003), which is a special case of the Jensen-Shannon divergence. We will use it for compositional data, since they also sum to 1. The main advantage of this newly proposed regression is that it handles zeros naturally and in addition it can lead to better fits as will be seen later.

The paper is structured as follows. Section 2 reviews some regression approaches for compositional data analysis and introduces the new regression model. Section 3 contains an example using real data and simulation studies to illustrate the performance of the newly proposed regression. Some Miscellanea and the Conclusion appear in Section 4 and 5 respectively.

2 Regression models

The most popular approach is to use the logistic normal distribution (Aitchison, 1982; 2003)

$$\log \left(\frac{\mathbf{y}_{-D}}{y_D} \right) = \left(\log \frac{y_1}{y_D}, \dots, \log \frac{y_d}{y_D} \right) \sim N_d \left(\mathbf{x}^T \mathbf{B}, \mathbf{\Sigma} \right),$$

where \mathbf{x}^T is a column vector of the design matrix \mathbf{X} , D is the number of components, $d = D - 1$ and

$$\mathbf{B} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_d) \quad \text{and} \quad \boldsymbol{\beta}_i = (\beta_{0i}, \beta_{1i}, \dots, \beta_{pi})^T, \quad i = 1, \dots, d.$$

are the regression coefficients and p is the number of independent variables. We will denote this regression as Aitchison regression.

A Dirichlet distribution (Gueorguieva et al., 2008) with the parameters linked to some covariates is another approach

$$\mathbf{y} \sim \text{Dir} \left(\frac{\phi}{1 + \sum_{j=2}^D e^{\mathbf{x}^T \boldsymbol{\beta}_j}}, \frac{\phi e^{\mathbf{x}^T \boldsymbol{\beta}_2}}{1 + \sum_{j=2}^D e^{\mathbf{x}^T \boldsymbol{\beta}_j}}, \dots, \frac{\phi e^{\mathbf{x}^T \boldsymbol{\beta}_d}}{1 + \sum_{j=2}^D e^{\mathbf{x}^T \boldsymbol{\beta}_j}} \right).$$

Note, that the precision parameter ϕ can also be linked to the independent variables via link function of course to ensure positivity of the estimated values. Maier (2014) has written an R package to fit Dirichlet regression models with and without covariates on ϕ .

Two approaches were examined in Murteira & Ramalho (2014) are the OLS regression

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^n [\mathbf{y}_i - \mathbf{f}_i(\boldsymbol{\beta}; \mathbf{x})] [\mathbf{y}_i - \mathbf{f}_i(\boldsymbol{\beta}; \mathbf{x})]$$

and the multinomial logit regression

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^n y_i \log \frac{y_i}{\mathbf{f}_i(\boldsymbol{\beta}; \mathbf{x})} = \max_{\boldsymbol{\beta}} \sum_{i=1}^n y_i \log \mathbf{f}_i(\boldsymbol{\beta}; \mathbf{x}),$$

where

$$\mathbf{f}_i(\boldsymbol{\beta}; \mathbf{x}) = \left(\frac{1}{\sum_{j=1}^D e^{\mathbf{x}_i^T \boldsymbol{\beta}_j}}, \frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}_2}}{\sum_{j=1}^D e^{\mathbf{x}_i^T \boldsymbol{\beta}_j}}, \dots, \frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}_d}}{\sum_{j=1}^D e^{\mathbf{x}_i^T \boldsymbol{\beta}_j}} \right).$$

Finally, the Kent distribution Kent (1982) was employed by Scealy & Welsh (2011) after the square root was applied component-wise to the data, thus ampping them to the unit (hyper)-sphere

$$\sqrt{\mathbf{y}} \sim \text{Kent}(\boldsymbol{\mu}(\boldsymbol{\beta}, \mathbf{x}), \boldsymbol{\Gamma}, \kappa, \boldsymbol{\delta}).$$

2.1 The ES-OV regression

We advocate that as a measure of the distance between two compositions we can use a special case of the Jensen-Shannon divergence

$$\text{ES-OV}(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^D \left(x_j \log \frac{2x_j}{x_j + y_j} + y_j \log \frac{2y_j}{x_j + y_j} \right), \quad (1)$$

where \mathbf{x} and $\mathbf{y} \in \mathbb{S}^d$. Endres & Schindelin (2003) and Österreicher & Vajda (2003) proved, independently, that (1) satisfies the triangular identity and thus it is a metric. The names ES-OV comes from the researchers' initials. In fact, (1) is the square of the metric, still a metric, and we will use this version.

The idea is simple and straightforward, minimization of the ES-OV metric between the observed and the fitted compositions with respect to the beta coefficients

$$\min_{\beta} \sum_{i=1}^D \left(\mathbf{y}_i \log \frac{2\mathbf{y}_i}{\mathbf{y}_i + \mathbf{f}_i(\beta; \mathbf{x})} + \mathbf{f}_i(\beta; \mathbf{x}) \log \frac{\mathbf{f}_i(\beta; \mathbf{x})}{\mathbf{y}_i + \mathbf{f}_i(\beta; \mathbf{x})} \right). \quad (2)$$

Below we summarise a few properties of the ES-OV metric and its associated regression model.

- The ES-OV metric belongs to the class of ϕ -divergences. Let $f(t) = t \log \frac{2t}{1+t} + \log \frac{2}{1+t}$, then $\text{ES-OV}(\mathbf{y}, \mathbf{z}) = \sum_{j=1}^D z_j f\left(\frac{y_j}{z_j}\right)$. Therefore, we can say that this regression falls within the minimum ϕ -divergence regression algorithms.
- A weighted version of the ES-OV regression (2), such as $\sum_{i=1}^D \left(\lambda \mathbf{y}_i \log \frac{2\mathbf{y}_i}{\mathbf{y}_i + \hat{\mathbf{y}}_i} + (1 - \lambda) \hat{\mathbf{y}}_i \log \frac{2\hat{\mathbf{y}}_i}{\mathbf{y}_i + \hat{\mathbf{y}}_i} \right)$ (Jensen-Shannon divergence) produces the best fits when $\lambda = 0$.
- A similar version of (2) is $\sum_{i=1}^D \left(y_i \log \frac{y_i}{\hat{y}_i} + \hat{y}_i \log \frac{\hat{y}_i}{y_i} \right)$, which gives nice results, but not as good (2).
- The ES-OV regression (2) can deal with zero values naturally as already mentioned in the Introduction.
- The ES-OV regression (2) can lead to better fits than the logistic normal.

The consistency and the asymptotic distribution of the regression parameters has not been studied. This is a task we have to do and we have confidence that the answers will be positive. As for the last property and in general the fit of regression models for compositional data we suggest the use of the Kulback-Leibler divergence (Kullback, 1997), which was also used as a measure of fit by Theil (1967)

$$\sum_{i=1}^n \text{KL}(\mathbf{y}_i, \hat{\mathbf{y}}_i) = \sum_{i=1}^n \mathbf{y}_i \log \frac{\mathbf{y}_i}{\hat{\mathbf{y}}_i}. \quad (3)$$

3 Example with real data and simulation studies

In both the example and the simulation studies we will compare the ES-OV regression with the Aitchison regression.

3.1 Real data analysis

In the first instance we will present an example with the Arctic lake (real) data we saw in Figure 1. Three ingredients, *sand*, *silt* and *clay* were measured at various depths (39 measurements) of an Arctic lake (Aitchison, 1986, p.359). The goal is to see how the composition in these three elements changes as the water depth increases and how good can our predictions be. We can see in Figure 2(a) that both methods fit the data well. As we move from left to right, we go from the surface to deeper levels of the lake. The composition of the samples has initially high percentage of sand, but as we move to the bottom of the lake, this percentage reduces, while the percentages of silt and clay increase. This makes absolute sense, since the percentages sum to 1, the derivatives, or the rates of change of the components must sum to zero. Figure 3 serves as an ancillary to Figure 2 for the readers who are not familiar with the ternary diagrams.

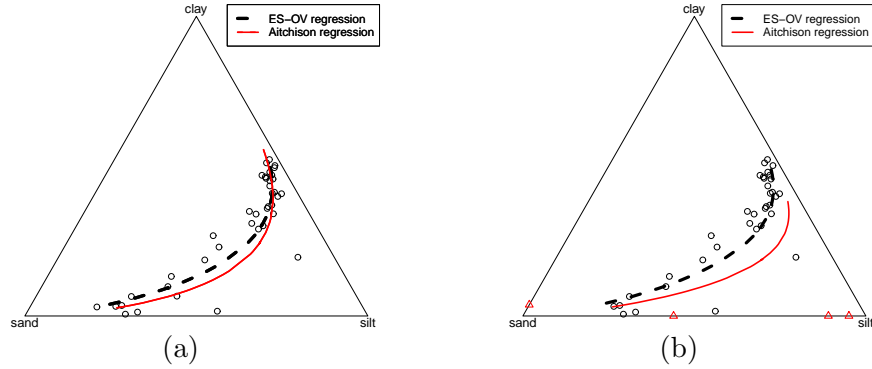


Figure 2: Regression lines for the Arctic lake data. The ternary diagram on the left contains the data as they are, whereas the right diagram has some zero values added.

We then make the values of four components equal to zero. In specific, the percentage of clay becomes zero for three observations and the percentage of silt becomes zero for one observation. In order to apply the logistic normal regression, imputation or zero replacement techniques (Martín-Fernández, 2012) are required. Templ et al. (2011) has created an R package (robCompositions) which performs zero value replacement for compositional data. Figure (2(b)) shows the Arctic lake data with the zero values (red triangles in the edges of the triangle). The Aitchison regression, applied to the new, zero value replaced data has been heavily affected by the zero values and the fitted values fall outside the bulk of the data. The ES-OV regression on the other hand seems unaffected.

3.2 Simulation studies

Simulation studies were conducted in the spirit of prediction performance. As we mentioned before, the asymptotic properties of the ES-OV regression, thus we cannot use it for inference

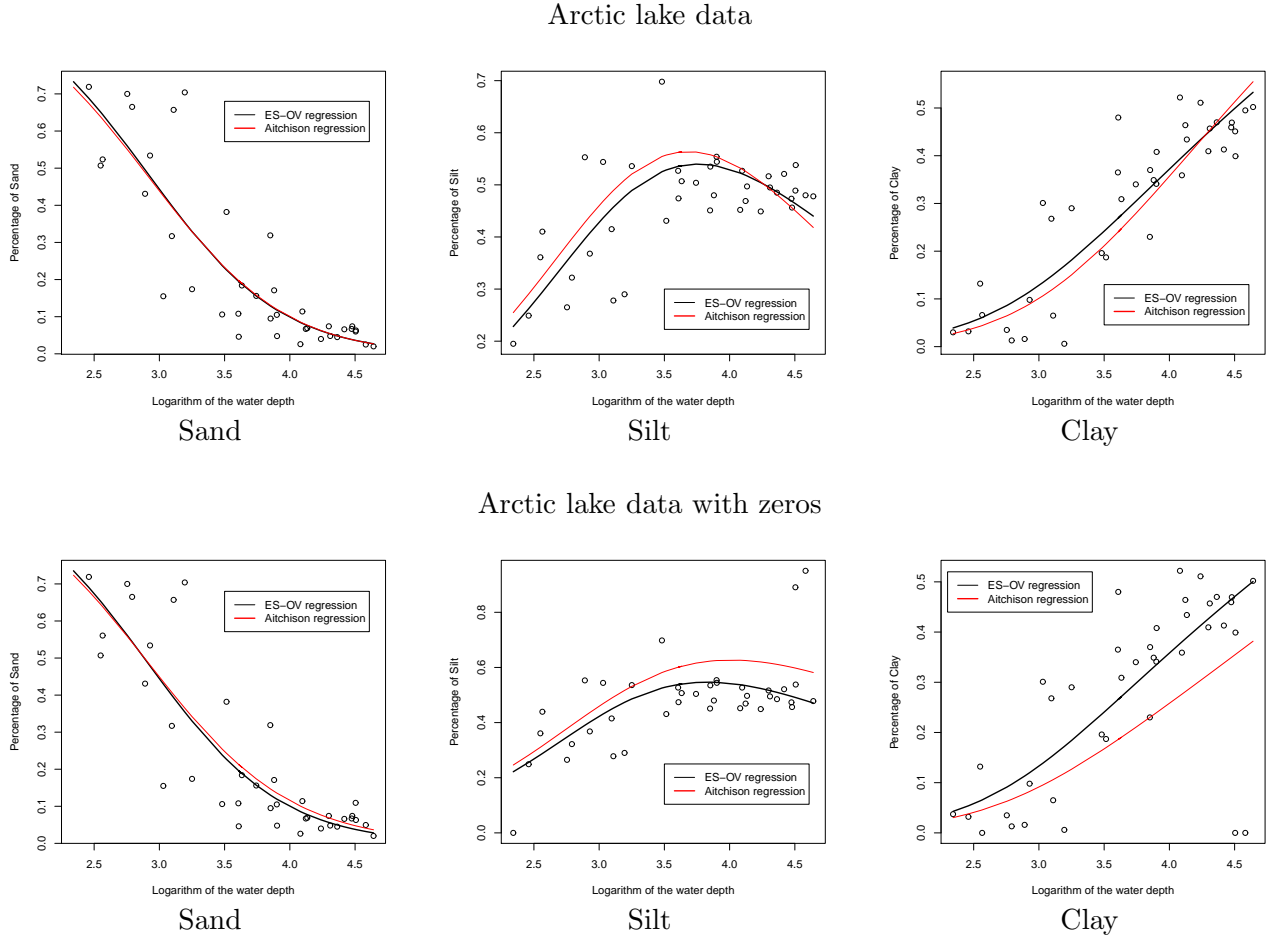


Figure 3: All graphs contain the logarithm of the water depth against the observed and the estimated percentages of the three components. The graphs in first row correspond to the observed Arctic lake data, while the graphs in the second row correspond to the data with some zeros added.

about the parameters. For this reason, we can only use it for prediction purposes. This is the current drawback of this regression.

We generated data from logistic normal regression models with various sample sizes ($n = 25, 50, 75, 100$) and components ($d = 6, 11, 16$). We repeated this scenario and made about half of the observations (simulating from a uniform variable) each time contain zero values. There were 2, 4 and 6 components with zero values when the simulated data consisted of 6, 11 and 16 components. For each case, we applied one-fold cross validation. That is, we remove one observation and estimate the regression parameters using the remaining data. Then we predict the value of the observation and calculate the Kullback-Leibler divergence of the true from the fitted value. This was repeated for all observations and in the end we summed all the divergences

$$\text{KL}(\mathbf{y}, \hat{\mathbf{y}}) = \sum_{i=1}^n \mathbf{y}_i \log \frac{\mathbf{y}_i}{\hat{\mathbf{y}}_{-i}},$$

where $\hat{\mathbf{y}}_{-i}$ is the i -th predicted compositional observation with the i -th observation having been excluded from the estimation of the regression parameters.

The one-fold cross validation procedure was repeated 200 iterations, due to limited compu-

tational sources, for each combination of sample size and number of components. In the end we averaged the Kulback-Leibler divergences

$$D = \frac{1}{200} \sum_{j=1}^{200} \text{KL}_j(\mathbf{y}, \hat{\mathbf{y}}). \quad (4)$$

The number 200 might seem small, but I think is large enough to make valid conclusions. The sample sizes considered are small, again due to limited computational resources. Even in this case, we believe we can extract some valid conclusions.

Compositional data with no zeros

The pseudo-code for the first set of the simulation studies is given below

Step 1. Generate n data from a multivariate normal regression model $\mathbf{Z} \sim N_p(\mathbf{B}\mathbf{X}, \mathbf{\Sigma})$, where \mathbf{X} is a design matrix with 2 independent variables. \mathbf{B} were chosen randomly from a standard normal distribution and $\mathbf{\Sigma}$ was a diagonal matrix with variances generated from an $\text{Exp}(1)$ distribution.

Step 2. Make the \mathbf{Z} compositional using

$$y_1 = \frac{1}{1 + \sum_{j=1}^p e^{z_j}}, \quad y_i = \frac{e^{z_i}}{1 + \sum_{j=1}^p e^{z_j}}, \quad \text{for } i = 2, \dots, p.$$

Step 3. Perform the ES-OV and Aitchison regressions and calculate (4) each time.

Step 4. Repeat Steps 1-3 for various small sample sizes $n = (25, 50, 75, 100)$ and number of variables $p = (5, 10, 15)$. Thus, we now have $D = (6, 11, 16)$ number of components and essentially 15, 30 and 45 beta parameters to estimate.

The results of this simulation study are presented in Table 1. We can see that the proportion of times, the Kulback-Lleibler divergence for the fitted values of the ES-OV regression is smaller than the fitted values of the Aitchison regression, grows large as the sample size increases, but when the sample size is equal to 100 decays. An explanation that can be given is due to random error. In addition, if we see Figure 4 we will see that the Kulback-Leibler divergences of the two regression models are close in general.

Sample sizes	Number of components		
	6	11	16
$n = 25$	0.195	0.105	0.055
$n = 50$	0.615	0.430	0.345
$n = 75$	0.820	0.705	0.565
$n = 100$	0.670	0.580	0.294

Table 1: Proportion of times the Kullback-Leibler divergence, of the true values from the fitted values, is smaller for the ES-OV than for the Aitchison regression. The data have no zero values.

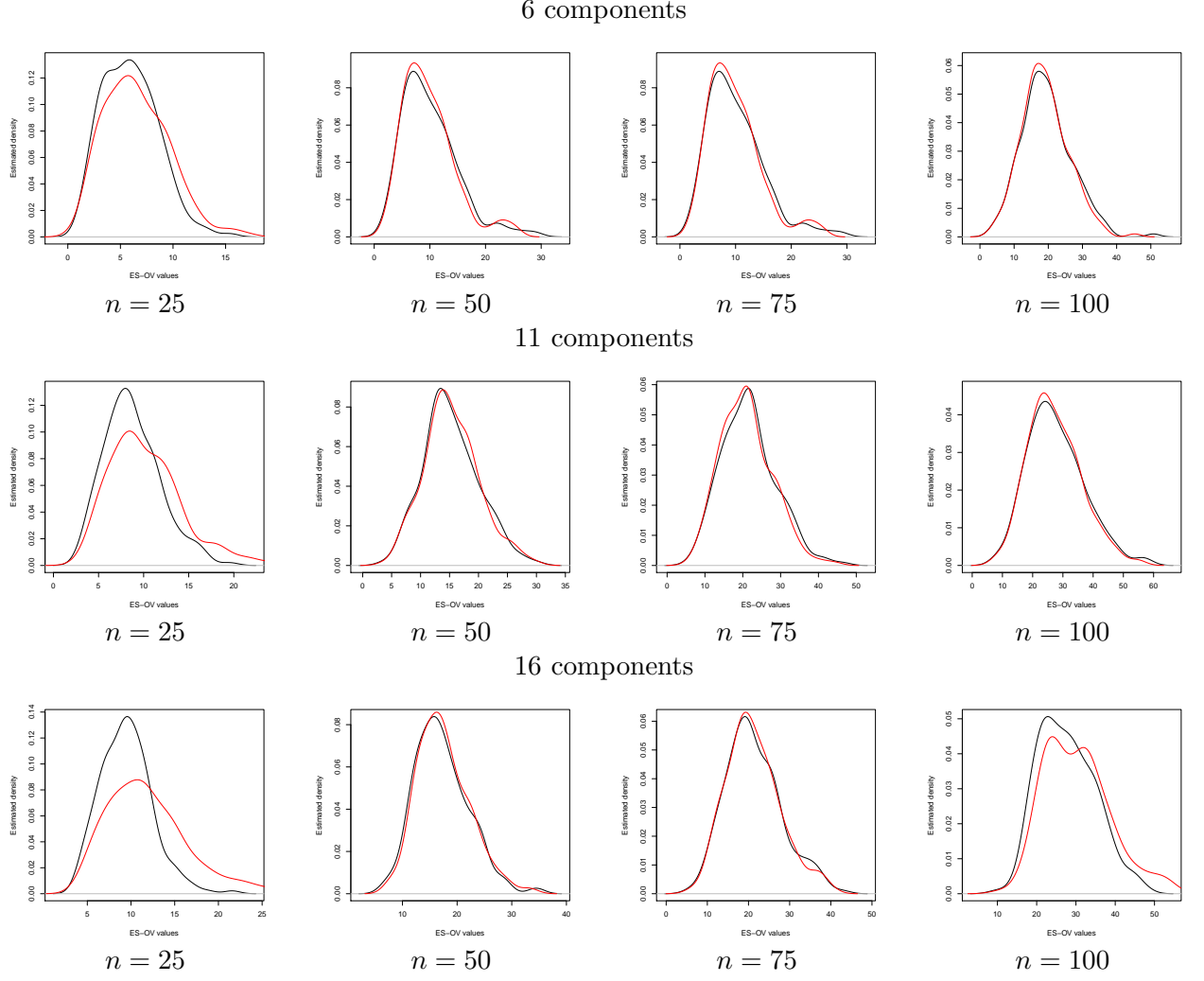


Figure 4: All graphs present the kernel estimated densities of the 200 Kullback-Leibler divergences of the ES-OV (red line) and the Aitchison regression (black line).

Compositional data with zeros

The calculations are the same as before but the scenario of the simulations has one modification: some values were set to zero

Step 1. Generate data from a multivariate normal regression model $\mathbf{Z} \sim N_p(\mathbf{B}\mathbf{X}, \mathbf{\Sigma})$, where \mathbf{X} is a design matrix with 2 independent variables. \mathbf{B} were chosen randomly from a standard normal distribution and $\mathbf{\Sigma}$ was a diagonal matrix with variances generated from an $\text{Exp}(1)$ distribution.

Step 2. Make the \mathbf{Z} compositional using

$$y_1 = \frac{1}{1 + \sum_{j=1}^p e^{z_j}}, \quad y_i = \frac{e^{z_i}}{1 + \sum_{j=1}^p e^{z_j}}, \quad \text{for } i = 2, \dots, p.$$

Step 3. Set about 50% (use a uniform distribution) of the elements of these components equal to zero and rescale these vectors so that their sum is again equal to 1.

Step 4. Perform the ES-OV regression and calculate (4).

Step 5. Use the R package *robCompositions* to replace the zero values and then apply the Aitchison regression and calculate (4).

Step 6. Repeat Steps 1-5 for various small sample sizes $n = (25, 50, 75, 100)$ and $p = (5, 10, 15)$. Thus, we now have $D = (6, 11, 16)$ number of components and essentially 15, 30 and 45 beta parameters to estimate.

Table 2 presents the the proportion of times the Kullback-Leibler divergence for the fitted values of the ES-OV regression is smaller than for the fitted values of the Aitchison regression. The results are clearer now, obviously the ES-Ov has managed to give better predictions, in terms of smaller Kullback-Leibler divergences.

Sample sizes	Number of components		
	6	11	16
$n = 25$	0.885	0.885	0.790
$n = 50$	0.975	0.995	0.995
$n = 75$	1.000	1.000	1.000
$n = 100$	1.000	1.000	0.990

Table 2: Proportion of times the Kullback-Leibler divergence, of the true values from the fitted values, is smaller for the ES-OV than for the Aitchison regression. The data have zero values.

4 Miscellanea

When $D = 2$, we end up with proportional data which can be analysed using a beta distribution. Kateri and Agresti (2010) used a ϕ -divergence regression model for binary responses. We can say, that in this cases the ES-OV is a special case of that model, since the ES-OV belongs to the class of ϕ -divergence statistics as we saw before.

Can we use another ϕ -divergence statistic for regression. Of course we can, the χ^2 distance and the Hellinger distance have been used and the results were very good. Again, the issue of consistency needs to be checked. Until this is done, they can be used for prediction purposes. In both of these cases, statistical properties about the parameters in the discrete case has been established. What about the compositional data case?

What happens when compositional data exist in the independent variables side? We request only for one condition, no zeros to exist in the independent variables compositional data. If this is satisfied, one can apply the centred log-ratio transformation (Aitchison, 2003) to the independent side compositional data

$$z_i = \log \frac{x_i}{g(\mathbf{x})}, \text{ for } i = 1, \dots, D,$$

where $g(\mathbf{x}) = \prod_{i=1}^{1/D}$ is the geometric mean of the components. Then, multiply the transformed data with the Helmert sub-matrix (Lee and Small, 1999), which is an orthonormal $D \times D$ matrix (Lancaster, 1965), whose first row, proportional to the vector of 1s, is deleted. This removes

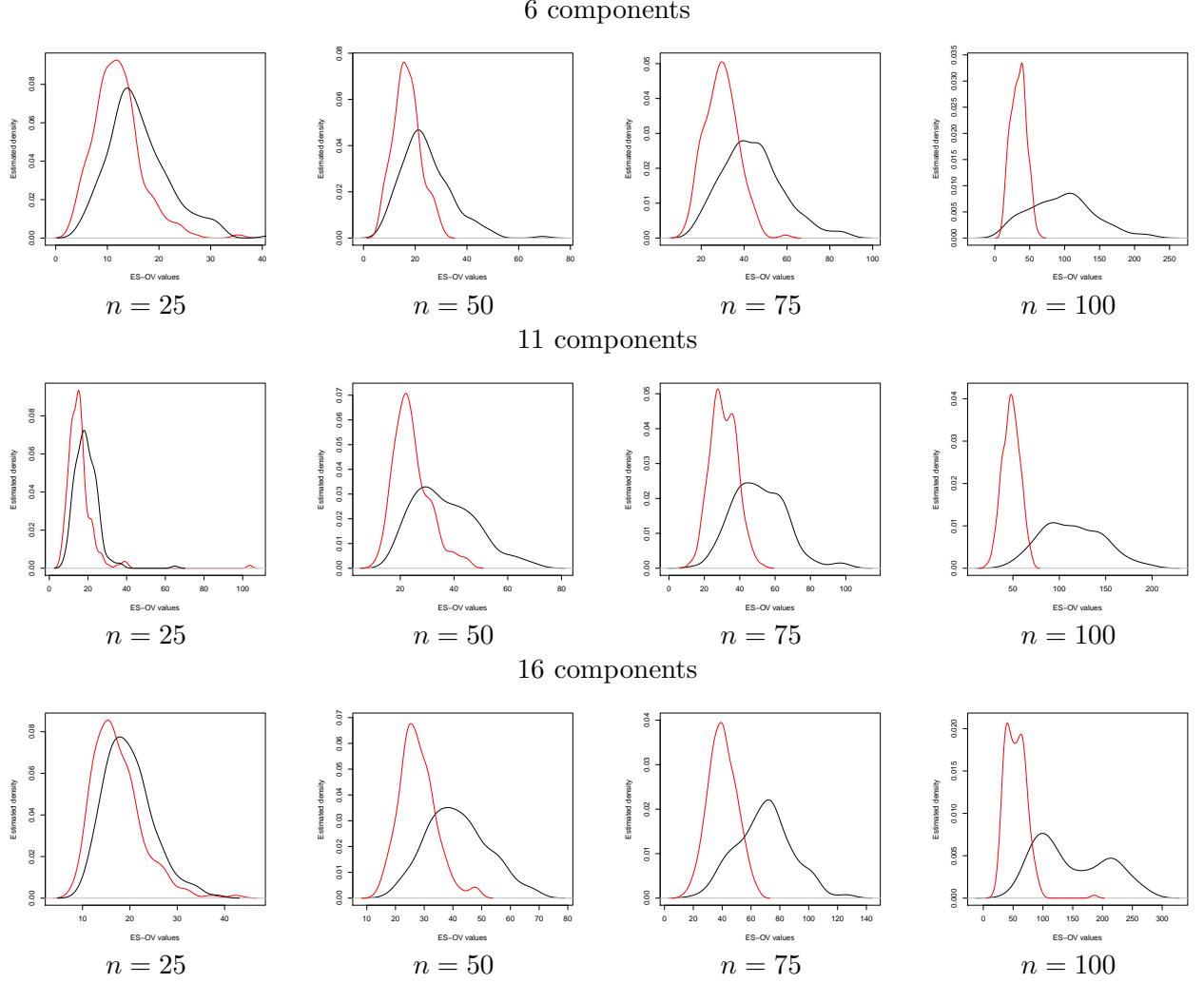


Figure 5: All graphs present the kernel estimated densities of the 200 Kullback-Leibler divergences of the ES-OV (red line) and the Aitchison regression (black line).

the singularity problem. Half of the job is done. Now, one can perform PCA on $\mathbf{Z}^T \mathbf{Z}$ and use the scores of the principal components, which is essentially a principal component regression (Jolliffe 2005). The number of principal components to be used can be tuned via cross validation by minimizing the Kullback-Leibler divergence.

5 Conclusions

We cannot say that we have proposed a new regression model that always outperforms the Aitchison regression. In fact we did not even see how it compares to the other methods mentioned in Section 2. We saw an example though where Aitchison regression fails. We also saw that in the case of many zeros, Aitchison regression is not very robust, but we believe this is a problem of the EM algorithm for the zero value replacements (Templ 2011) and not of the regression itself. For these cases, the ES-OV regression does better than Aitchison regression. The goal was not to find a model that outperforms always all the other existing ones, but to propose one which is equally good or better in some cases. In the case of no zero values,

Aitchison regression seems to do a better job, but in the case of many zeros, ES-OV regression is suggested.

In addition, since we have not examined the asymptotic properties of the ES-OV regression, this method is currently advisable to be used for prediction purposes. So, if the interest is to estimate as accurately as possible the compositions of new observations, given a set of observed data containing lot of zeros, we suggest the use of the ES-OV regression.

Acknowledgements

The author would like to express his gratitude to Aziz Alenazi (PhD student in statistics at the university of Sheffield) and to Theo Kypraios (Assistant professor in statistics at the university of Nottingham) for their help with the computations.

Appendix: R codes

```
### Aitchison regression
```

```
compreg <- function(y, x) {
  ## y is dependent variable, the compositional data
  ## x is the independent variable(s)
  ## the next two commands make sure the data are matrices
  y <- as.matrix(y)  ## makes sure y is a matrix
  y <- y/rowSums(y)  ## makes sure y is compositional data
  z <- log( y[, -1] / y[, 1] )  ## additive log-ratio (alr) transformation
  ## with the first component being the base
  n <- nrow(z)  ## sample size
  d <- ncol(z)  ## dimensionality of z
  p <- ncol(x)  ## dimensionality of x
  X <- as.matrix( cbind(1, x) )  ## the design matrix
  beta <- solve(t(X) %*% X) %*% t(X) %*% z  ## the parameters
  est1 <- X %*% beta  ## fitted values
  est2 <- cbind(1, exp(est1))
  est <- est2 / rowSums(est2)
  list(beta = beta, fitted = est)
}
```

```
### ES-OV regression
```

```
esov.compreg <- function(y, x) {
  ## y is dependent variable, the compositional data
  ## x is the independent variable(s)
```

```

y <- as.matrix(y)
y <- y/rowSums(y) ## makes sure y is compositional data
x <- as.matrix( cbind(1, x) )
d <- ncol(y) - 1 ## dimensionality of the simplex
n <- nrow(y) ## sample size
z <- list(y = y, x = x)
reg <- function(para, z = z){
  y <- z$y ; x <- z$x
  be <- matrix(para,byrow = T,ncol = d)
  mu1 <- cbind(1, exp(x %*% be))
  mu <- mu1/rowSums(mu1)
  M <- (mu + y)/2
  f <- sum( y * log(y / M) + mu * log(mu / M), na.rm = T )
  f
}
## the next lines minimize the reg function and obtain the estimated betas
ini <- as.vector( t( coef( lm(y[, -1] ~ x[, -1]) ) ) ) ## initial values
val <- NULL
qa <- nlm(reg, ini, z = z)
val[1] <- qa$minimum
qa <- nlm(reg, qa$estimate, z = z)
val[2] <- qa$minimum
i <- 2
while (val[i-1] - val[i] > 0.00001) {
  i <- i + 1
  qa <- nlm(reg, qa$estimate, z = z)
  val[i] <- qa$minimum
}
val <- min(val)
beta <- matrix(qa$estimate, byrow = T, ncol = d)
mu1 <- cbind(1, exp(x %*% beta))
mu <- mu1 / rowSums(mu1)
list(beta = beta, val = val, fitted = mu)
}

```

REFERENCES

- Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society. Series B* **44**, 139-177.
- Aitchison, J. (2003). *The statistical analysis of compositional data*, New Jersey: Reprinted by The Blackburn Press.

- Endres, D. M. and Schindelin, J. E. (2003). A new metric for probability distributions. *Information Theory, IEEE Transactions on* **49**, 1858-1860.
- Gueorguieva, R., Rosenheck, R., and Zelteman, D. (2008). Dirichlet component regression and its applications to psychiatric data. *Computational statistics & data analysis* **52**, 5344-5355.
- Jolliffe, I. T. (2005). *Principal component analysis*, New York: Springer-Verlag.
- Kateri, M. and Agresti, A. (2010). A generalized regression model for a binary response. *Statistics & Probability Letters* **80**, 89-95.
- Kent, J. T. (1982). The Fisher-Bingham distribution on the sphere. *Journal of the Royal Statistical Society. Series B* **44**, 71-80.
- Kullback, S. (1997). *Information theory and statistics*, New York: Dover Publications.
- Lancaster, H. (1965). The Helmert matrices. *American Mathematical Monthly* **72**, 4-12.
- Le, H. and Small, C. G. (1999). Multidimensional scaling of simplex shapes. *Pattern Recognition* **32**, 1601-1613.
- Maier, M. J. (2014). DirichletReg: Dirichlet Regression in R. *R package version 0.5-2*.
- Martín-Fernández, J., Hron, K., Templ, M., Filzmoser, P., and Palarea-Albaladejo, J. (2012). Model-based replacement of rounded zeros in compositional data: Classical and robust approaches. *Computational Statistics & Data Analysis* **56**, 2688-2704.
- Murteira, J. M. and Ramalho, J. J. (2014). Regression analysis of multivariate fractional data. *Econometric Reviews (ahead-of-print)* 1-38.
- Neocleous, T., Aitken, C., and Zadora, G. (2011). Transformations for compositional data with zeros with an application to forensic evidence evaluation. *Chemometrics and Intelligent Laboratory Systems* **109**, 77-85.
- Österreicher, F. and Vajda, I. (2003). A new class of metric divergences on probability spaces and its applicability in statistics. *Annals of the Institute of Statistical Mathematics* **55**, 639-653.
- Scealy, J. L. and Welsh, A. H. (2011). Regression for compositional data by using distributions defined on the hypersphere. *Journal of the Royal Statistical Society. Series B* **73**, 351-375.
- Stephens, M. A. (1982). Use of the von Mises distribution to analyse continuous proportions. *Biometrika* **69**, 197-203.
- Templ, M., Hron, K., and Filzmoser, P. (2011). robCompositions: Robust estimation for compositional data. *R package version 0.8-4*.
- Theil, H. (1967). *Economics and information theory*. Amsterdam: North-Holland publishing company.